The Role of Prefrontal Cortex Context Maintenance and Updating in Predictive Sentence Processing

Yasemin Gokcen, Rachel Ryskin, and David Noelle
Department of Cognitive and Information Science
University of California, Merced

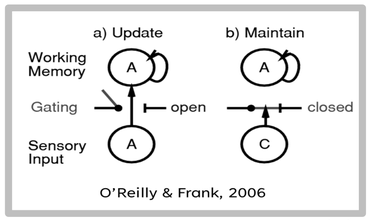
BACKGROUND

Language

- During language comprehension, humans routinely use information from preceding context to predict upcoming input (Ryskin & Nieuwland, 2023).
- The memory representation of the context is noisy (Hahn et al., 2022).
- Surprisal represents the expectations of a word given context (Levy, 2008).
  - Lower surprisal = better performance

Working memory & cognitive control

- Computational models of working memory (WM) mechanisms associated with prefrontal cortex (PFC) use *gating* to capture the ability to maintain and update contextual/task-relevant information in memory (O'Reilly et al., 2002).
- The anatomy and physiology of the PFC is thought to support this mechanism (Noelle, 2012; Kriete et al., 2013).



- Memory is updated when gate opens to relevant information.
- Memory is maintained when gate is closed to irrelevant information.
- Models/the brain learn when the gate should be open vs. closed

QUESTION

To what degree are the gating mechanisms of prefrontal cortex used to maintain context for brain systems shown to be sensitive to word surprisal?

GENERAL APPROACH AND METHODOLOGY

Compare performance across language models with different architectures & their ability to predict neural measures of real-time language processing.

Language Models:

- Recurrent Neural Network (RNN) and Long Short-Term Memory models (van Schjindel & Linzen, 2018) both trained on the Wikitext 2 dataset and tested on Natural Stories corpus (Futrell et al., 2021).
  - **LSTM has PFC-like gating mechanisms**, RNN does not (Hochreiter & Schmidhuber, 1997)

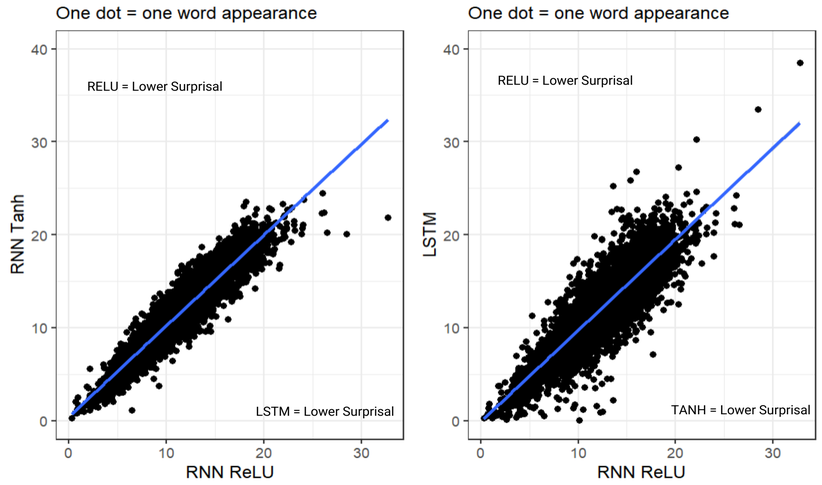
Neural activity:

- EEG passive listening task (+ comprehension questions)
  - Natural Stories corpus
  - 10 stories/narratives (~5 min each) with low frequency words/structures
  - Data collection in progress

COMPARING SURPRISAL ACROSS LANGUAGE MODELS

Correlations between word-by-word surprisal values:

- High correlation (r = .949) when comparing between RNN models (Tanh vs. ReLU activation functions) and between LSTM and RNN (r = .893).
- Trained 5 iterations of each model with different random initializations to ensure robustness.
- Words for which LSTM surprisal is lower than RNN surprisal may indicate sentence contexts where gating plays an important role due to WM demands.



WM DEMANDS AND RELATIVE SURPRISAL ACROSS MODELS

Working Memory Demands of Context

Previous theories posit measures of the WM requirements of processing a sentence, based on sentence contents.

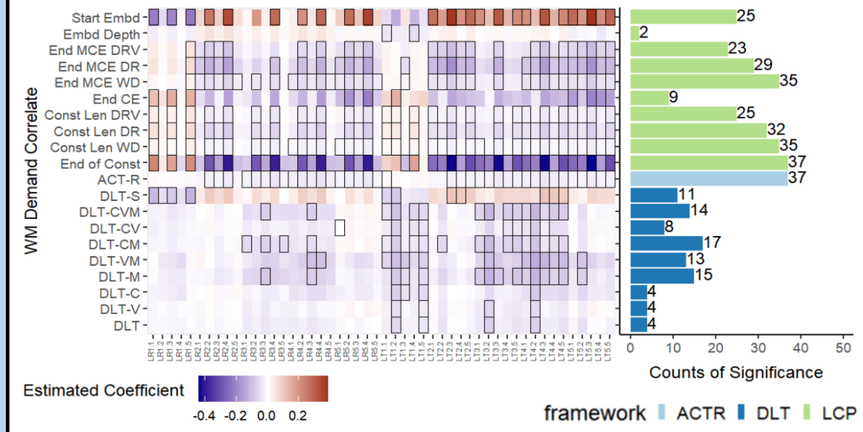
We examined 3 frameworks:

- Dependency Locality Theory (Gibson, 2000) (DLT)
  - Greater distance between words in a dependency incurs greater WM cost
  - For example: How many intervening words are between a subject and a verb within a sentence?
- Adaptive Character of Thought—Rational (ACT-R) (Anderson et al., 2004)
  - A cognitive architecture
  - Provides a measure of WM retrieval costs and does not consider storage/maintenance
- Left-Corner Parsing (Lewis & Vasishth, 2005)
  - Constituents (i.e., phrases) and embeddings (i.e., phrase within a phrase)
  - WM demand increases as simultaneous processing of embeddings and constituents within a sentence increases
  - Markers for phrase endings, distance metrics, and memory depth tracking

WM DEMANDS AND RELATIVE SURPRISAL ACROSS MODELS

Residuals from each linear model predicting LSTM surprisal from RNN surprisal were regressed onto WM demand measures

Note: Tiles with outline indicate a regression coefficient that is statistically significant based on a Bonferroni-adjusted α (DLT α=0.0055; ACT-R α=0.05; LCP α=0.01).



- Correlate measures of WM demand with RNN vs. LSTM surprisal residuals:
  - Negative → LSTM outperforms RNN more as WM demand increases
  - Positive → RNN outperforms LSTM more as WM demand increases

- Across frameworks, most measures of WM demand are negatively associated with residuals suggesting that, as WM demand increases, LSTM models are more likely to outperform RNN at next-word prediction.
- Modest evidence that WM demands posited by DLT are related to effects of model architectural differences on prediction (i.e., presence of gating in LSTMs)
- Stronger evidence that WM demands posited by LCP (except Start Embedding) and ACT-R are related to effects of model architectural differences on prediction (i.e., presence of gating in LSTMs) (though only one measure for ACT-R).

Preliminary conclusions:

- PFC-like gating enables LSTM models to predict upcoming language input more efficiently than similar-performing RNN models without gating when the WM demands of the sentence context are high (e.g., center embedding, long constituent)
- PFC gating mechanism may critically support prediction from context during sentence processing in humans

FUTURE WORK

Further investigation of properties of sentence context associated with gating mechanisms (updating, maintenance)

Predicting EEG responses during story listening from language model metrics (e.g., surprisal, residual)